Why Economists Ignored the Method of Least Squares*

Jinbang Kim
University of California

The method of least squares was proposed in 1805 and soon became a standard tool in astronomy and geodesy. But it is not until the mid-twentieth century that the same method could be found in the ordinary toolbox of economists. The delayed adoption is explained by the intended scope of application of the method in relation to the nature of economics as the nineteenth century economists understood it. Given a parametric relation between a few variables and their inaccurate measures, scientists employed the method to obtain the "most probable" value of parameters. By contrast, it was believed that an economic relation contained so many variables and would never be deduced in a parametric form. As long as economists adhered to the Newtonian deduction, they would not adopt the method of least squares which they considered only as a way of combining inaccurate measures.

I. Introduction

It is not until the mid-twenieth century that econometrics could be found in the ordinary toolbox of economists.[1] The enterprise of using statistical methods to find a numerical relation between variables, however, is much older in other disciplines. In particular, the method of least squares, which is still one of the most important procedures in econometrics, was proposed and became a standard tool in astronomy and geodesy almost two centuries ago. It is thus not surprising that the list of Merriman (1877) contained the titles of 408 papers and books relating to the method even after numerous

*For useful suggestions, the author would like to thank an anonymous referee and participants in the Economic Thought Workshop at Duke University.

[1]Textbooks in econometrics became available in the early 1950's Tinter (1952) and Klein (1953) were among them.

[Seoul Journal of Economics 1991, Vol. 4, No. 4]

practical papers had been left unnoticed.

The slow diffusion of techique across disciplinary boundary is not unusual in intellectual history. But three points should be noted about the case at hand. First, the method of least squares is not much involved. Its principle is simple to understand and its algorithm is easy to implement. Second, it seems that Jevons (1871) considered to employ the method to ascertain a demand curve.[2] But his contemporaries and immediate followers showed little interest in such a line of researches. Third, the method has suddenly become a popular tool in economics after the long delay. A glance at academic journals published in the 1950's and 1960's will suffice to confirm this movement.

These facts lead to various questions, one of which I attempt to answer in the present paper: What barred the nineteenth century economists from adopting the method of least squares? Granted that many of the economists were aware of the method and technically capable of using it, I suspect that there would have been some conceptual barriers delaying its application to economic investigations. I will thus focus on how the economists understood the method in relation to economic problems.

II. The Method of Least Squares As Scientists Understood It

Let us start with the characterization of the method of least squares by its inventor, a French scientist named Adrien M. Legendre.[3] He explained its intended scope of application as follows.[4]

In the majority of investigations in which the problem is to get

[2]For the details of Jevons's applications, see Kim (1992), Aldrich (1987) and Stigler (1986, 1992).

[3]In 1805 Legendre proposed the method of least squares in an appendix to his essay about a comet's orbit. Four years after its first appearance in print, a German scientist named Carl F. Gauss claimed that he had been using the same method since 1795. The claim apparently provoked Legendre, and resulted in a protracted priority dispute between the two men. See Plackett (1972).

[4]The quotations are from an English translation by Ruger and Walker in Smith (1929). The pages, however, refer to that of Legendre (1805). Here the difference between Legendre's and today's terminology should be noted. For instance, what Legendre called "the known coefficients" and denoted by a, b, c were in fact the measurements of variables; a modern statistician may denote such quantities by x, y, z with a subscript like i. Bue Legendre himself used x, y, z to denote "the unknowns" which a modern statistician would denote by α, β, τ. "The number of equations" is also equivalent to "the number of observations," and "a system of equations" should be understood accordingly.

> from measures given by observation the most exact results which they can furnish, there almost always arises a system of equations of the form
>
> $$E = a + bx + cy + fz + etc.$$
>
> in which a, b, c, f, *etc.* are the known coefficients which vary from one equation to another, and x, y, z, *etc.* are the unknowns which must be determined in accordance with the condition that the value of E shall for each equation be reduced to a quantity which is either zero or very small.
>
> If there are the same number of equations as unknowns x, y, z, *etc.*, there is no difficulty in determining the unknowns, and the error can be made absolutely zero. But more often the number of equations is greater than that of the unknowns, and it is impossible to do away with all the errors. (Legendre 1805, p. 72)

We can understand from this quotation that the method of least squares presupposed a parametric relationship between variables. Given such a relationship and some measures of the variables, the method was applied to deduce the value of parameters, especially in the case where the number of observations was greater than the number of parameters to be deduced.

An illustration will help to make the point more clear.[5] Since Newton suggested that the earth had the shape of an ellipsoid with a bulging at the equator, scientists including Boscovich engaged themselves in ascertaining the exact relation between arc length and latitude. Of course, the relation could be derived from the hypothesis and approximated by the following linear-in-parameters equation.[6]:

$$a = z + y\sin^2\theta$$

where a is the length of 1° of latitude centered at latitude θ, and z and y are unknown coefficients. Further, the value of coefficients could be deduced from the measures of a at two different locations. The problem, however, was that there were more than two measures of them available and each of the pairs yielded a different result. Indeed Boscovich deduced ten different values for y from ten different combinations out of five measures of a and θ. It is in "a situation like this, which is the usual thing in physical and astronomical problems," that Legendre proposed to use the method of least

[5]This illustration is based on Stigler (1986).
[6]The notation is the same as Boscovich's.

squares (Legendre 1805, p. 72). To repeat, the method was intended to determine the value of coefficients in an equation based on conflicting observations, where the equation was derived from a theory or hypothesis.

There is another point we should keep in mind. Boscovich or his contemporaries was not surprised at all to find that each set of the measures yielded a different value for the parameters. The scientists understood that the inconsistency might have resulted from the errors in measurement. This view led Legendre to start his exposition of the method of least squares with an equation containing a term denoted by E. Legendre must have included the term to represent the measurement errors. Note also that he located the term on the opposite side of other terms, which suggests that E was included to represent the measurement errors in *any or all of the variables.*[7] De Morgan, a prominent mathematician who taught Jevons at the University College, made the point more explicit when he discussed the method of least squares: "Let us now suppose two results of observation, say that we wish to know the fraction which A is of B, where both A and B are subject to errors..." (1838, p. 156).

This being the view about the relation between measured values of variables, the nineteenth century scientists also believed that the measurement errors would cancel each other if the measures were properly combined. Hence, the scientists looked for a proper way of combining the conflicting measures instead of giving up their efforts to determine the values of unknown coefficients.[8] The scientists would have been happier if they had a general principle rather than a case-by-case or ad hoc solution. They would have been even happier if the general principle had been easy to apply.[9] It is such a

[7]This interpretation might lead a student of modern statistics to wonder: How, if any, did Legendre deal with the complication resulting from the measurement errors in a dependent variable? As the issue is too involved to address in this paper, I will just suggest the following. The question above will make sense to a "classical" statistician. But Legendre and his contemporaries were living in a different tradition. For instance, Laplace (1812) had no difficulty in computing the "inverse probability" of an least-squares estimate no matter which variable had been measured with errors.

[8]Though this may be obvious to today's scientists, it was not so to many of the mid-eighteenth century scientists like L. Euler. He contended that we can conclude nothing from such conflicting measures. See Stigler (1986, pp. 25-30).

[9]Boscovich (1755) himself proposed a general principle. He determined the values for z and y by taking the arithmetic average of the ten values deduced from the ten combinations out of five observations on α and θ. This average method, however, would have required enormous computation if the number of observations had not been so limited. For instance, if Mayer (1750) had adopted the average method to determine the moon's

general and easy way of combining the inaccurate measures that Legendre proposed in 1805.

According to the principle proposed by Legendre, z and y should be estimated by the values which minimize the sum of E^2's, or

$$\sum_{i=1}^{5}(a_i - z - y\sin^2\theta_i)^2$$

Legendre also demonstrated that such values are to be found by solving only two equations which he called the equations of minimum (les équations du minimum)[10]:

$$0 = \sum_{i=1}^{5}(a_i - z - y\sin^2\theta_i)$$
$$0 = \sum_{i=1}^{5}\sin^2\theta_i(a_i - z - y\sin^2\theta_i)$$

That is, we have to solve only as many equations as unknowns no matter how many measurements we have. Legendre thus claimed that "of all the principles which can be proposed ... there is none more general, more exact, and more easy of application" than his (Lengendre 1805, p. 75).[11]

Now let us summarize what we can make of the above explications. First, the method of least squares was intended to apply to a relationship of which the form and the relevant variables were deduced from a theory or a set of theories. Second, the method was based on the view that measurement error is responsible for the inconsistency between a theory-suggested relationship and measures. Third, the method was regarded as a general and easy way of combining inaccurate measures.

III. Inexactness and Complexity of Economic Relations

We have so far explicated how scientists understood the method

equator based on 27 observations, he should have solved a three-equation system as many as 2925 times. But he had a different idea, and employed a method which depended on an *ad hoc* rule. For the details, see Stigler (1986).

[10]In the second half of his appendix on the method of least squares, Legendre actually showed how the method could be applied to the Boscovich's problem. Legendre, however, started with a different, though equivalent, equation and used different measures. Furthermore, his application was not so straightforward as that described in this paper.

[11]Unlike two other merits of Legendre's principle, its exactness might not be immediately apparent to his contemporaries. Even the meaning of exactness might be ambiguous to them. Nevertheless, Legendre attempted no explication or justification of the claim. It is Laplace (1812) who gave an explicit meaning of the "exactness" and proved that the method of least squares would yield the most "exact" as well as the most 'probable" estimate. The proof was an improvement upon that by Gauss (1809).

of least squares. It would be safe to claim that the nineteenth century economists would have had the same view, if any, about the scope of application of the method. Then I shall contend in this section the following: the nineteenth century economists considered the method of least squares useless for their investigations because the economic relations were so complicated and insusceptible of exactness.

It seems that the first part of the proposition above is difficult to establish. Since few economists had attempted to adopt the method to quantify an economic relation, the economists did not have to explicate their view about the applicability of the method to economic investigations.[12] However, we can still notice the general attitude of economists against any use of statistics as a means of ascertaining economic laws. For instance, Cairnes (1875) repeatedly pointed out the "utter fertility" and the "necessary impotence" of induction in economic inquiry. He also took a case to make the point more clear. The case concerned a tabular statement about the relations of the harvest and price of corn, which Davenant attributed to Gregory King.[13] It was evident to Cairnes that "no reliance can be placed on the accuracy of such calculations" (Carines 1875, p. 116), no matter how the table had been produced. Interestingly, when Moore (1911, 1914) employed the method of least squares as an essential tool for his study of demand, the response of Marshall was more blunt. He wrote Edgeworth that "Moore is a nightmare to me."[14]

But we should not be surprised that the nineteenth century economists were unwilling to adopt the method of least squares. Given the nature of economics as they understood it, it would have been evident to the economists that the method was inapplicable to their

[12]As was said earlier, Jevons (1871) seems to have used the method to estimate a demand curve. However, so far as I know, his practice constituted an exception and other economists simply ignored it.

[13]To reproduce the table,

Defect		Above the Common Rate.
1 Tenth	raises the price	3 Tenths
2 Tenths		8 Tenths
3 Tenths		1.6 Tenths
4 Tenths		2.8 Tenths
5 Tenths		4.5 Tenths

[14]Marshall to Edgeworth, 5 June 1912, Henry L. Moore Papers, Rare Book and Manuscript Library, Columbia University. The statistical works of Moore are briefly examined in Epstein (1987), Mirowski (1989) and Morgan (1991).

investigations. Most of all, economics differs from astronomy or geodesy as the former seldom suggests a *particular form of relation* between a *limited number of variables*. In economics a variable is usually explained as an *unspecific function* of an almost *infinite number of variables*. In addition, many of the relevant variables defy quantification or measurement. That is, economics does not yield an equation to which the method of least squares was originally intended to apply. The nineteenth century economists, who must have shared the inventor's view about the method, could hardly take the invention of the method to mean "an important addition to the economist's box of tools."[15]

Admittedly, it cannot be more than a conjecture that the nineteenth century economists decided not to adopt the method of least squares after considering the nature of economics in relation to the intended scope of application of the method. But so far as the nature of economics is concerned, the economists could not have been more empathetic. For example, J. B. Say stressed the number of variables which would determine the price of wine.

> We may, for example, know that for any given year the price of wine will infallibly depend upon the quanitity to be sold, compared with the extent of the demand. But if we are desirous of submitting these two data to mathematical calculation, their ultimate elements must be decomposed... Hence, it is not only necessary to determine what will be the product of the succeeding vintage... but the quality it will possess, the quantity remaining on hand of the preceding vintage, the amount of capital that will be at the disposal of the dealers... All these data, and probably many others besides, must be accurately appreciated, solely to determine the *quantity* to be put in *circulation*; itself but one of the elements of *price*. To determine the *quantity* to be *demanded*... we must also know the former stock on hand and the tastes and means of the consumers... (Say 1826, p. 26-7)

Besides the number of relevant variables, Say was also concerned about the problem of quantification and specification.[16]

> I suppress an infinite number of less important considerations, more or less affecting the solution of the problem; for I

[15]Schumpeter revealed his "feeling of surprise" at the neglect of the method of squares by economists for such a long time (Schumpeter 1954, p. 525).

[16]According to Ménard (1980), Say was opposed to the statistical approach mainly because of his belief that statistical data are necessarily transitory and dangerously linked to state power.

> question whether any individual... would even venture to attempt this, not only on account of the numerous data, but in consequence of the difficulty of characterizing them with anything like precision, and of combining their separate influence. (Say 1826, p. 27)

The unattainability of specific or exact laws was further elaborated by Cairnes. Like most scientists of the Newtonian era, he believed that a scientific law should be deduced from some fundamental principles and economics should not differ in this respect. That is, an economic law should be deduced from the principles of human nature and the physical conditions. Such a law could be of a specific form if both the mental and physical principles were of a specific form. But it is never the case.

> For, although the general character of [the mental] principles may be ascertained, and although when stated with sufficient precision they may be made the basis of important deductions, yet they do not, from the nature of the case, admit of being weighed and measured like the elements and force of the material world: they are therefore not susceptible of arithmetic or mathematical expression; and hence it happens that, in speculating on results which depend on the positive or relative strength of such principles, perfect precision, numerical accuracy, is not attainable. Political economy seems on this account necessarily excluded from the domain of exact science. (Cairnes 1875, p. 109)

In his letter to Moore, Marshall also noted two of the three characterícts of economics, namely the complexity of economic relations and the difficulty in quantifying economic entities. The 1911 book of Moore, Marshall said, seemed to proceed on line which he himself "had deliberately decided not to follow years ago." He then gave two reasons for the decision. One was basically that most economic events are associated with a large number of causes, and the other was that many of the cause "have refused as yet to be tabulated statistically."

We now understand that at least one obstacle could have been resolved by giving a new meaning (and, perhaps, a new name) to what was called an error term.[17] That is, Say would not have worried about "an infinite number of less important considerations" if he had noticed that an error term could be reinterpreted to capture them.

[17]It is interesting that we have been using the name "error term" even after it stopped representing the measurement error.

That is, he might have been able to concentrate on relationship between a limited number of variables while having an error term represent the deviations due to omitted variables. This semantical reinterpretation of error term, however, was not immediate at all. It seems to be the case even for W. S. Jevons, a representative nineteenth century economist of quantitative mind and probably the first to apply the method of least squares to economic data. Like Legendre and De Morgan, he took the method as the best way to combine inaccurate measures and nothing else. In the *Principles of Science,* for instance. Jevons explained the method in the following words:

> If the result of each observation [on a, b and c] gives an equation between [x and y] of the form
>
> $$ax + by = c$$
>
> then, *if the observations were free from error,* we should need only two observations giving two equations; but for the attainment of greater accuracy, we may take many observations... The mean equations having been computed, their solution... gives the most probable values of the unknown quantities. (Jevons 1887, p. 394; emphasis added)

Probably, the reinterpretation of error term might have been hindered by a dominant view of the world in the eighteenth and nineteenth century. Scientists including economists had long objected to any suggestion that the world might be governed by probabilistic laws. Therefore, even the effects of less important elements should be deterministic. Scientists could not think of representing the effects by the same law that they applied to the measurement errors.

For this and other reasons, economists waited for Yule (1897a, 1897b, 1899) to give a wholly new meaning to the method of least squares and thus to the error term. But since Yule's renovation of the method of least squares is not the main issue of the present paper, I will not discuss it beyond the following. Yule seems to be the first to view the method of least squares as a way of retrieving the curve of conditional mean, or what is called the regression line.[18] That is, he supposed a probabilistic relation between vari-

[18]The term "regression" was introduced by F. Galton to imply, for instance, that "the mean height of the sons of parents of a given type is nearer to the mean height of the general population then their parents' height is" (Yule 1897a, p. 814n). We now use the same term whether the curve of conditional mean is of a slope lower than 45 degree or not.

ables, rather than such a relation only between their measures. This supposition opened the path to reinterpretation of error term. The effects of omitted variables could be represented by a probabilistic law and assigned the statistically same status at the measurement errors. "A certain proportion of... change," said Yule, "must be of a purely chance character, i.e., due to unspecifiable causes like the changes in numbers thrown in casts of dice." It would not be exaggeration to call this view revolutionary to economists. Moore might have adopted such a view as he acquainted himself with Pearson and Yule at the Galton laboratory. But T. Haavelmo was the economist who led the revolution in economics while attempting "to supply a theoretical foundation for the analysis of interrelations between economic variables" (Haavelmo 1944, p. 3).

IV. Conclusion

The applicability of the method of least squares to economic studies could not be so obvious to the nineteenth century economists as it is to today's students of economics. On one hand, the nineteenth century economists must have shared the inventor's view about the scope of application of the method, i.e., the view that the method is a way of deducing the value of parameters from a parametric relationship between variables and inaccurate measure of the variables. On the other hand, the economists were concerned about the problem inherent in economics; most economic events are associated with a large number of causes, whereas some of the causes defy quantification. Further, economics seldom assigns a particular form to the association.[19] This view and concern seem to have delayed the diffusion of the method of least squares into economics.

How then did economists come to adopt the method of least squares? It has been suggested that Galton expedited the adoption by giving a new meaning to the method in connection with a probabilistic view of the world. But further study is needed to claim the importance of Yule in the history of econometrics.

[19]In this respect, I disagree with Stigler (1986, p. 359) who concludes as follows. "It was not that theory was lacking: By the middle of the nineteenth century several economists had given mathematical expression to the theories of Adam Smith and David Ricardo." I would note that mathematical expression was not enough; a parametric relationship was still lacking in economics.

References

Aldrich, J. "Jevons as Statistician: the Role of Probability." *Manchester School* 55 (June 1987) : 253-7.

Cairnes, J. E. *The Character and Logical Method of Political Economy.* 2nd edition, London: Macmillian, 1875.

Epstein, R. J. *A History of Econometrics,* North-Holland, 1987.

Gauss, C. F. *Theoria motus corporum coelestium,* 1809. Translated as *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections.* trans. C. H. Davis, Boston: Little Brown, 1963.

Harvey, G. "On the Method of Minimum Squares, Employed in the Reduction of Experiments, Being a Translation of the Appendix to an Essay of Legendre's, Entitled 'Nouvelles Méthodes pour la Determination des Orbites des Cometes,' with Remarks." *The Edinburgh Philosophical Journal* 7 (1822) : 292-301.

Haavelmo, T. *The Probability Approach in Econometrics.* Supplement to *Econometrica* 12 (July 1944).

Jevons, W. S. *The Theory of Political Economy.* London: Macmillian, 1871.

____________. *The Principles of Science.* 2nd edition, London: Macmillian, 1887.

Kim, J. "Jevons's Curve Fitting: A Reconstruction." Unpublished paper, 1992.

Klein, L. R. *A Textbook of Econometrics.* New York: Row, Peterson and Company, 1953.

Laplace, P. M. *Théorie Analytique des Probabilités,* 1812. Translated as *A Philosophical Essay on Probabilities.* trans. F. Truscott, and F. Emory, New York: Dover Publications, 1951.

Legendre, A. M. *Nouvelles Méthodes pour la Determination des Orbites des Cometes.* Paris: Courcier, 1805.

Ménard, C. "Three Forms of Resistance to Statistics: Say, Cournot, Warlas." *History of Political Economy* 12 (Winter 1980) : 524-41.

Merriman, M. "A List of Writings Relating to the Method of Least Squares with Historical and Critical Notes." *Transactions of the Connecticut Academy of Arts and Sciences* 4 (1877) : 151-232.

Mirowski, P. "The Probabilistic Counter-Revolution, or How Stochastic Concepts Came to Neoclassical Economic Theory." *Oxford Economic Papers* 41 (January 1989) : 217-35.

Moore, H. *Laws of Wages : An Essay in Statistical Economics.* 1911. Reprinted, New York: Augustus M. Kelly Publishers, 1967.

____________. *Economic Cycles : Their Law and Cause,* 1914. Reprinted, New York: Augustus M. Kelly Publishers, 1967.

Morgan, M. *The History of Econometric Ideas.* Cambridge University Press, 1990.

Placket, R. L. "The Discovery of the Method of Least Squares." *Biometrica* 59 (1972) : 239-51.

Porter, T. M. *The Rise of Statistical Thinking, 1820-1900.* Princeton University Press, 1986.

Say, J. B. *Traité d'Économie Politique*, 1826. Tranlated as *A Treatise on Political Economy.* trans. C. R. Prinsep, and C. C. Biddle, Reprinted, New York, 1964.

Schumpeter, J. A. *History of Economic Analysis.* Oxford University Press, 1954.

Smith, D. E. *A Source Book in Mathematics.* New York, 1929.

Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press, 1986.

________. "Stigler as Statistician." *Manchester School* 50 (1986): 354-65.

________. "Jevons on the King-Davenant Law of Demand: Another Example of Non-least Squares." Unpublished paper, 1992.

Tinter, G. *Econometrics.* New York: John Wiley & Sons, 1952.

Yule, G. U. "On the Theory of Correlation." *Journal of Royal Statistical Society* 60 (1897): 812-54. (a)

________. "On the Significance of Bravis' Formulae for Regression, &c. in the Case of Skew Correlation." *Proceedings of the Royal Society of London* 60 (1897): 477-89. (b)

________. "An Investigation into the Causes of Changes in Pauperism in England, Chiefly during the Last Two Intercensal Decades, Part I." *Journal of the Royal Statistical Society* 62 (1899): 249-86.